Lending Club Project Part 2

Team 5

1. *Created dummy variable for term length called long\_term*
2. *Created dummy variables for each grade called A, B, C, D, E, and F*
3. What is the mean and standard deviation of the interest rate charged to the customers?
   1. *Mean: 12.62*
   2. *Standard Deviation: 5.26*
4. Run the correlations between variables you think might be important to predict the interest rate. Note down which variables are highly correlated with interest rate. Also note which of these variables that you think would predict interest rate are correlated with one another. Fun fact: if you are running a simple regression model, the square of the correlation is the regression coefficient.
   1. *As a precursor, we selected 15 variables to determine which are highly correlated with predicting what affects the interest rate. During our R&D, found 10**specific variables with a correlation coefficient about 22%, our marker for high correlation. These variables helped us see a direct correlation to our interest rate, even though there were a couple of variables highly correlated with one another we still had enough to prove a fit model.*
   2. *Here are our variables used to help, All\_util, long\_term (dummy variable for length of loan), and A-F grading from LC (dummy variables), bc\_open\_to\_buy, and total\_rev\_hi\_lim.*
   3. *We start by introducing two variables that were found to be highly correlated to one another; Tot\_cur\_bal and Tot\_hi\_cred\_lim having a 98% direct correlation to each other.*  *Neither are strongly correlated to the interest rate. However, we did find Bc\_open\_to\_buy and Total\_rev\_hi\_lim have a strong positive correlation of 69% to each other and are both highly correlated, by our standard, to the interest rate.*
5. In your report, submit briefly your correlation matrix from #4, and summarize your findings.
   1. *In our findings of tot\_cur\_bal & tot\_hi\_cred\_lim, and Bc\_open\_to\_buy & Total\_rev\_hi\_lim due to each having a high correlation to one another, each independent variable in this subset with high correlation to one another we understand should not be used in our model as determinates.*
   2. *Bc\_open\_to\_buy is the total open to buy on revolving bank cards has a negative correlation of 31%, suggesting the higher the bc\_open\_to\_buy, the lower the interest rate. Total\_rev\_hi\_lim has a significant negative correlation to the interest rate, suggesting the higher the total revolving high credit limit the lower the interest rate. There was a strong negative correlation of roughly 72% between grade A and interest rate, leading us to believe higher graded borrowers receive lower rates. The correlation between the other grades are the interest rate generally followed this idea as well, especially D (54%), E (44%), and F(32%).*
   3. *It is important to note that we generally chose variables that had at least a 22% positive or negative correlation to interest rate to be considered highly significant and all grades given by the Lending Club. The grades are assigned based on the history of the borrower’s credit. They have a particularly high correlation and importance to the Lending Club when assigning interest rates.*
   4. *The long\_term variable is a dummy variable referring to the length of the loan (36 months =0, 60 months =1). There is a positive correlation between the length of the loan the interest rate, suggesting the more months the loan is spread over, the higher the interest rate. All\_util variable, refers to balance to credit on all trades has a positive correlation of 30% to the interest rate. This positive correlation indicates that as All\_util increase, so does the interest rate.*